Dossier n°7 : Exemple d'organisation et d'étude d'une série statistique. Détermination, comparaison, utilisation de mesures de tendance centrale (paramètres de position) et de mesures de dispersion (paramètres de dispersion). Regroupement en classes. Représentations graphiques usuelles.

Rédigé par Cécile COURTOIS, le 26 août 2003
cecile-courtois@wanadoo.fr

I Situation par rapport aux programmes.

L'étude des séries statistiques commence au collège.

En classe de cinquième, on introduit les notions de classes, d'effectifs et de fréquences ainsi que les premières représentations graphiques de séries.

En quatrième, les élèves voient les notions d'effectifs et de fréquences cumulées ainsi que les moyennes pondérées.

En troisième, on introduit les notions de médiane et d'étendue.

Toutes les notions vues au collège sont reprises en seconde.

En Première S et ES, les élèves étudient la variance et l'écart type, les intervalles interquartiles et les diagrammes en boîtes.

Je choisis donc de situer ce dossier en Première S et ES.

II Commentaires généraux.

II.1 A propos du sujet.

La place des statistiques dans la vie quotidienne est devenue très importante et pour cela, il est primordial d'apprendre aux élèves à les manipuler de façon raisonnée.

Face à une grande quantité de données chiffrées, il faut savoir réagir. Pour cela, il faut organiser ces données de façon simple et logique pour pouvoir par la suite les interpréter. C'est alors qu'on introduit diverses caractéristiques d'une série statistique (position ou dispersion) afin de pouvoir par la suite émettre un avis critique sur les données présentées.

Le but de ce dossier est donc de faire une synthèse des méthodes d'organisation et d'étude de séries statistiques.

II.2 A propos des exercices.

J'ai donc choisi, pour illustrer ce dossier, de vous présenter trois exercices :

- l'exercice n°1 propose l'étude classique d'une série statistique présentée sous forme discrète, puis regroupée en classes ;
- l'exercice n°2 propose l'étude de deux séries statistiques puis de la série constituée du regroupement des deux précédentes ;
- l'exercice n°3 propose la comparaison de deux séries statistiques.

Rappelons les divers paramètres d'une série statistique. Pour cela, on considère une série statistique prenant pour valeurs $(x_i)_{1\leq i \leq p}$ avec pour effectifs $(n_i)_{1\leq i \leq p}$ respectivement et d'effectif total n.

Paramètres de position :

- *Moyenne :*

La moyenne d'une série discrète est le nombre réel noté $\bar{x}$ et égal à $\frac{n_1x_1 + \ldots + n_px_p}{n}$.

Dans le cas d'un caractère continu ou discret regroupé en classes, on calcule la moyenne en choisissant comme valeurs du caractère les centres des classes.

- *Médiane :*

On suppose les valeurs du caractère ordonnées par ordre croissant.

Si n = 2k alors la médiane est la demi-somme des termes de rang k et k+1 dans cette série ordonnée.

Si n = 2k+1 alors la médiane est la valeur du terme de rang k+1 dans cette série ordonnée.

Si la série est continue, la médiane est le réel qui partage la population en deux parties de même effectif.

Paramètres de dispersion :

- *Ecart type :*

L'écart type est le nombre réel s = $\sqrt{V}$ où V = $\frac{1}{n}\sum_{i=1}^{p} n_i x_i^2 - \bar{x}^2$.

Commentaire : On peut montrer que :

- au moins 68% de l'effectif a une valeur du caractère qui appartient à l'intervalle $[\bar{x}-s ; \bar{x}+s]$;
- au moins 75% de l'effectif a une valeur du caractère qui appartient à l'intervalle $[\bar{x}-2s ; \bar{x}+2s]$;
- au moins 88% de l'effectif a une valeur du caractère qui appartient à l'intervalle $[\bar{x}-3s ; \bar{x}+3s]$;

Si l'écart type d'une série est faible, les intervalles précédents sont de faibles amplitudes et la moyenne est fiable et fournit un bon renseignement sur le caractère étudié.

- *Quartiles :*

Le premier quartile Q_1 est la plus petite valeur du caractère telle qu'au moins 25% des termes de la série aient une valeur du caractère qui lui soit inférieure ou égale.

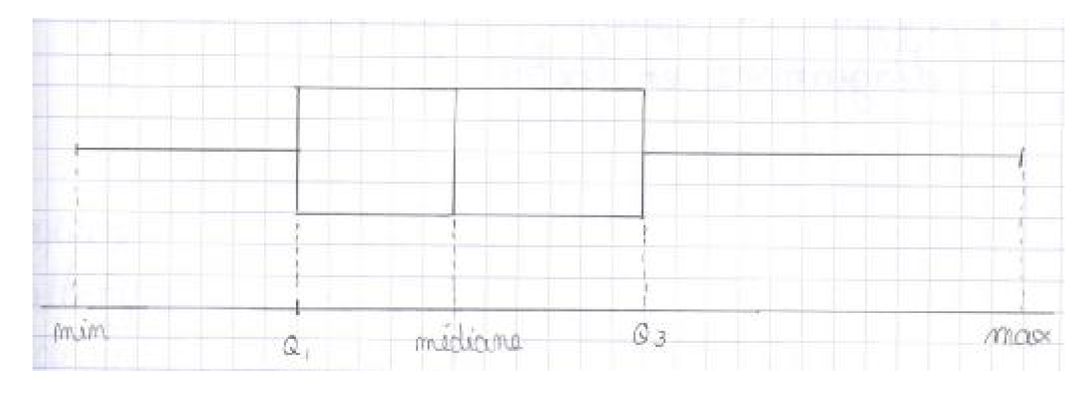
Le troisième quartile Q_3 est la plus petite valeur du caractère telle qu'au moins 75% des termes de la série aient une valeur du caractère qui lui soit inférieure ou égale.

L'intervalle interquartile est l'intervalle $[Q_1 ;Q_3]$. Le nombre I = Q_3-Q_1 est l'interquartile.

Diagrammes en boîtes :

On construit un diagramme en boîtes de la façon suivante :

- les valeurs du caractère sont repérées sur un axe (vertical ou horizontal) ;
- on place sur cet axe le minimum et le maximum de la série, le premier quartile et le troisième quartile ainsi que la médiane ;
- on construit alors un rectangle (la boîte) parallèlement à l'axe, dont la longueur est l'interquartile.



Dans les exercices proposant des séries statistiques de grand effectif, on utilisera la calculatrice et les outils qu'elle propose pour le calcul des paramètres.

III Présentation des exercices.

III.1 Exercice n°1.

<u>But :</u> Etudier la série des masses limites de rupture en charge pour 48 fils.

<u>Objectif :</u> Comparer les paramètres d'une série discrète et de cette même série regroupée en classes.

<u>Outils :</u>
- Moyenne ;
- Ecart type ;
- Médiane ;
- Interquartile ;
- Histogrammes ;
- Polygones des effectifs ;
- Diagramme en boîtes.

III.2 Exercice n°2.

<u>But :</u> Etudier la répartition des salaires nets des hommes, des femmes et de l'ensemble d'une entreprise.

<u>Objectif :</u> Déduire de certaines paramètres de deux séries, ceux de la série constituée par leur regroupement sans en connaître les valeurs.

<u>Outils :</u>
- Fréquences ;
- Polygone des fréquences ;
- Diagrammes en boîtes ;
- Moyenne ;
- Ecart type.

III.3 Exercice n°3.

<u>But :</u> Etudier les séries des âges de décès des hommes et des femmes d'une même ville pour une année.

<u>Objectif :</u> Comparer deux séries.

<u>Outils :</u>
- Moyenne ;
- Ecart type ;
- Médiane ;
- Quartiles ;
- Diagrammes en boites ;
- Coefficient de variation :$C_v = \frac{s}{\bar{x}}$.

<u>Commentaire :</u>

Le coefficient de variation relativise la valeur de l'écart type par rapport à la moyenne.

Ainsi, il permet de comparer entre elles des séries dont les valeurs ont des ordres de grandeur différents. En effet, si un caractère prend des valeurs très grandes par rapport à un autre, l'écart type du premier sera certainement supérieur à celui du second, mais cela ne veut pas dire que la dispersion du premier caractère est plus grande que celle du second.

IV Enoncés et références des exercices.

IV.1 Exercice n°1 (n°73 p 191, Belin 1ère S 2001).

Pour étudier la résistance d'un métal, on choisit au hasard un échantillon de 48 fils de même diamètre et on note les masses limites de rupture en charge pour chacun de ces fils. Les résultats, en grammes, sont donnés par le tableau :

713	908	846	703	772	910	895	917	765	926	893	798	792	865	895	792	821	890	882
706	852	794	790	931	925	711	703	903	903	724	869	931	852	895	852	704	805	921
759	939	759	888	785	787	903	893	787	917									

1. a) Calculer la masse limite moyenne en grammes du métal et l'écart type de la série statistique.
   b) Déterminer l'écart interquartile et la médiane.
2. a) Regrouper les valeurs et donner les effectifs des classes [700 ;720[, [720 ;760[, [760 ;800[, [800 ; 840[, [840 ;880[, [880 ;900[, [900 ;920[et [920 ;940[.
   b) Tracer l'histogramme et le polygone des effectifs.
   c) Calculer la moyenne et l'écart type.
   d) Calculer l'écart interquartile et la médiane.
3. Comparer les résultats des questions 1 et 2.
4. En reprenant les données du tableau, représenter le diagramme en boîtes de la série.

IV.2 Exercice n°2 (TD1 p 192, Transmath 1ère S 2001).

Dans une entreprise de 130 salariés, la répartition des salaires nets (hors primes et hors dirigeants) est la suivante, en millier d'euros :

Classe	[1 ;1,25[	[1,25 ;1,5[	[1,5 ;1,75[	[1,75 ;2[	[2 ;2,25[	[2,25 ;2,5[	[2,5 ;2,75[	[2,75 ;3[
Effectif masculin	12	21	17	10	4	3	2	1
Effectif féminin	18	22	13	4	2	1	0	0

1. Pour chaque série, déterminer les fréquences cumulées croissantes et construire les polygones sur le même graphique.
2. En vous aidant de ce graphique, construire un diagramme superposé en boîtes des deux séries.
3. Pour chaque série, calculer la moyenne, la moyenne des carrés et l'écart type.
4. On souhaite calculer la moyenne et l'écart type de la série des salaires de l'ensemble des salariés, sans faire le tableau de regroupement des deux séries.
   a) Déduire de la question 3 la moyenne de l'ensemble des salaires. En déduire aussi la somme des carrés des salaires de la série regroupée puis la moyenne de ses carrés.
   b) Calculer l'écart type de la série.

IV.3 Exercice n°3 (n°41 p 200, Transmath 1ère S 2001).

On a rassemblé pour une ville donnée les âges de décès des hommes et des femmes pour une année.

Hommes : 25-65-79-59-58-77-73-75-68-66-75-71-63-89-66-77-84-84-79-77-59-60-81-45-63-75-65-84-31-67-82-74-78-86-76-56-59-56-78-84-62-48-79-41-60-41-5-61-73-76-63-68-52-68-87-62-72-52-71-79-81-80-80-58-86-75-83-63-80-59-84-64-64-79-79-63-63-79-71-72-34-43-59-53-68-65-89-54-59-59-83-77-77

Femmes : 94-82-63-56-90-88-72-76-94-89-67-88-82-82-61-97-88-68-89-91-56-85-77-68-81-88-45-73-97-78-83-74-91-80-76-80-95-62-98-73-66-90-71-73-66-55-53-94-57-69-62-81-80-90-81-43-76-91-88-87-73-80-87-87-84-64-78-77-64-54-81-88-84-85-74-44-84-97-54-76-87-59-80-72-90-90-70-87

1. Calculer pour chaque série :
   a) la moyenne et l'écart type ;
   b) la médiane et les quartiles.

2. Etablir les diagrammes superposés en boîtes des deux séries.

3. Calculer le coefficient de variation C_V pour chaque série égal à $\frac{s}{\overline{x}}$ où s est son écart-type et $\overline{x}$ sa moyenne. Quelle remarque peut-on faire ?

V Commentaires.

a) On remarque que l'énoncé du dossier est très long et très complet : il faudra impérativement que tous les éléments proposés dans cet intitulé soient présents dans les exercices. J'espère n'en avoir oublié aucun mais il est très souvent possible d'ajouter à un exercice des questions permettant de le compléter et de répondre entièrement au problème posé.

b) Il est stipulé dans le dossier proposé par le jury qu'au moins un des exercices devra faire appel à la calculatrice. Il est également demandé de mettre en évidence, à travers ce dossier, la dimension civique des mathématiques.

c) Je tenais à vous signaler une erreur sur la TI-89, ce qui va vous permettre de garder tous vos cheveux sur la tête.

En effet, j'ai remarqué, en effectuant certains calculs sur des séries statistiques, que la formule utilisée par la TI-89 pour calculer l'écart type était fausse. Pour vous en rendre compte, il vous suffit de vous reporter au manuel de la calculatrice, dans la liste de toutes les fonctions et de regarder à écart-type (p 204) pour constater que la formule utilisée par la calculatrice est la suivante :

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{p} n_i x_i^2 - \bar{x}^2}$$

ce qui peut donner des résultats tout à fait aberrant.

Voici comment corriger cette erreur. Vous entrez votre série statistique dans le tableur puis vous en calculez la moyenne. Vous retournez ensuite dans le tableur et vous ajoutez une (p+1)ème valeur à votre série : $x_{p+1} = \bar{x}$ et $n_{p+1} = 1$. Ainsi, la moyenne de votre série reste inchangée et le calcul de l'écart type est juste (vérification facile).

Attention : Ceci permet de corriger le calcul de l'écart type et n'influe pas sur celui de la moyenne mais pourrait changer certains autres calculs effectués par la calculatrice sur votre série.

d) Les dossiers sur les statistiques demandent beaucoup de calculs, il faut donc garder du temps lors de la préparation pour les effectuer.

e) Plusieurs modèles de calculatrice proposent notamment le tracé des diagrammes en boîtes et de beaucoup d'autres types de représentations graphiques des séries statistiques.